

CONTEMPORARY REVIEW

The Impact of Novel Assessment Methodologies in Toxicology on Green Chemistry and Chemical Alternatives

Ivan Rusyn^{*,1} and Nigel Greene[†]^{*}Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, Texas 77843; and[†]Predictive Compound Safety and ADME, AstraZeneca Pharmaceuticals LP, Waltham, Massachusetts 02451¹To whom correspondence should be addressed at Department of Veterinary Integrative Biosciences, Texas A&M University, 4458 TAMU, College Station, TX 77843. Fax: 979-458-9866; E-mail: irusyn@tamu.edu.**ABSTRACT**

The field of experimental toxicology is rapidly advancing by incorporating novel techniques and methods that provide a much more granular view into the mechanisms of potential adverse effects of chemical exposures on human health. The data from various *in vitro* assays and computational models are useful not only for increasing confidence in hazard and risk decisions, but also are enabling better, faster and cheaper assessment of a greater number of compounds, mixtures, and complex products. This is of special value to the field of green chemistry where design of new materials or alternative uses of existing ones is driven, at least in part, by considerations of safety. This article reviews the state of the science and decision-making in scenarios when little to no data may be available to draw conclusions about which choice in green chemistry is “safer.” It is clear that there is no “one size fits all” solution and multiple data streams need to be weighed in making a decision. Moreover, the overall level of familiarity of the decision-makers and scientists alike with new assessment methodologies, their validity, value and limitations is evolving. Thus, while the “impact” of the new developments in toxicology on the field of green chemistry is great already, it is premature to conclude that the data from new assessment methodologies have been widely accepted yet.

Key words: alternatives to animal testing; *in vitro* and alternatives, predictive toxicology; systems biology; methods.

Considerations that go into “green chemistry” decisions, more precisely defined as assessments of chemical alternatives, inevitably must include concerns regarding human health. The traditional hazard identification step still applies because the adverse health endpoints and the sources of data are largely identical. Assessments of chemical alternatives, however, typically are based on a comparative approach do not include dose-response or weight-of-evidence mechanistic evaluations. Assessments of chemical alternatives “is different from a safety assessment, where the primary goal is to ensure that exposure is below a prescribed standard; different from risk assessment, where risk

associated with a given level of exposure is calculated” (National Research Council, 2014).

Hazard identification relies on multiple data streams. These include human epidemiologic or experimental studies, animal bioassays, and increasingly numerous *in vitro* and *in silico* novel assessment methodologies (Rusyn and Daston, 2010). The National Academies reports *Toxicity Testing in the 21st Century (TT21C): A Vision and a Strategy* (National Research Council, 2007) and *Using 21st Century Science to Improve Risk-Based Evaluations* (National Academies of Sciences, 2017) endorsed the use of novel assessment methodologies in decision-making. Similarly, advances in chemistry and material sciences create new

opportunities for more green chemistry and use of alternative materials in products and processes.

Although it is clear that assessments of alternatives will benefit from more information, it is less obvious how exactly data derived from novel assessment methodologies—can fit into decision-making, or whether “more information” equates to “better information”. Identification of the needed data types and the tools to be used to analyze these data, as well as a clear understanding of how different data streams will be used in the assessment of alternatives should be clearly identified by the user during the early scoping stage of the process. No blueprints and only some examples exist of how *in vitro* data can be incorporated into benchmarking approaches to assess health hazards. Experimental and computational tools to generate and integrate novel data streams into the process of assessing alternatives is assuming growing importance; however, the uptake of these data for decision-making on environmental and commodity chemicals is lagging, especially in the context of green chemistry and use of chemical alternatives.

This review summarizes the findings of the National Academies' Committee on the Design and Evaluation of Safer Chemical Substitutions with respect to the use of *in vitro* data and *in silico* models for chemical alternatives assessments (National Research Council, 2014). The committee concluded that these data are most useful as primary data in human health hazard assessments and to fill data gaps. Importantly, it was concluded that “the approaches used for novel data streams, especially the broad range of end points provided by high-throughput assays, may be less amenable to a formal classification scheme”. Instead, it was envisioned that “user-defined decision rules and principles will likely guide incorporation of these data into the alternatives assessment process. As a result, the expert, judgment-guided discussions with regulatory bodies may need to remain flexible depending on the type of chemical alternatives assessment.”

CONCEPTUAL PROPOSALS FOR THE USE OF *IN VITRO* DATA AND *IN SILICO* MODELS TO ASSESS HUMAN HEALTH HAZARDS

One risk assessment approach aimed at deriving a level of human intake or exposure to a chemical that is perceived to be of negligible risk, despite the absence of chemical-specific toxicity data, is the Threshold of Toxicological Concern (TTC) (Cramer *et al.*, 1978). The classification of chemical substances of concern had been originally developed to qualitatively assess the risk of low-level substances in the diet, but is now used frequently to determine whether a comprehensive risk assessment is required for a broad range of chemicals. It is used as a means of waiving testing based on knowledge of exposure limits. A number of decision trees (Roberts *et al.*, 2015), databases (Tluczkiewicz *et al.*, 2011), and software packages (Bhatia *et al.*, 2015) have been developed that can aid TTC assessments for new chemicals (Roberts *et al.*, 2015). The application of the TTC approach to a broader universe of chemicals and routes of exposure has been the main focus of recent research, along with a number of proposals for the revision of the decision tree to derive a TTC (Tluczkiewicz *et al.*, 2016).

In addition to the existing approaches like TTC, it is now feasible to obtain complex data on hundreds, if not thousands, of chemicals (Collins *et al.*, 2008; Kavlock *et al.*, 2012; Tice *et al.*, 2013). The field of environmental health sciences is nearly unanimous in the opinion that human and environmental

health decisions will be made with new data; however, there are differences in the view on when and how a transition from traditional *in vivo* data to new assessment methodologies should occur and what type of a decision (ranking/prioritization for further traditional testing, making choices about alternatives, etc.) is most appropriate (Andersen and Krewski, 2009, 2010; Krewski *et al.*, 2011).

Furthermore, *in silico* tools for predicting adverse effects have existed for several decades. Most of these methods have focused on hazard identification. The concept of chemical similarity has been used to develop a variety of methods to predict chemical-induced responses based only on chemical structure, or on a combination of chemical and biological information (Greene and Song, 2011; Grimm *et al.*, 2016, 2017; Low *et al.*, 2014; Rusyn *et al.*, 2012). Indeed, read-across analyses (Enoch *et al.*, 2008; Hewitt *et al.*, 2010; Voutchkova *et al.*, 2010) are easily adoptable for assessments of chemical alternatives.

The following are examples of approaches that make use of *in vitro* data in human health assessments that also are useful with respect to assessments of alternatives:

1. Crump *et al.* (2010) argued that *in vitro* data can be used in ways similar to current process of risk assessment, except that additional safety factors may need to be applied to account for extrapolation from *in vitro* to *in vivo*. The authors also argued that pathways-based models of toxicity (National Research Council, 2007) may not be useful for quantitative decision-making. The statistical variability inherent in complex models may hinder their ultimate utility for estimating small changes in response. Furthermore, such models involve empirical modeling of dose-responses.
2. Judson *et al.* (2011) proposed using *in vitro* data and pharmacokinetic models, coupled with estimates of population variability and uncertainty, to estimate the human dose at which a chemical may significantly alter a biological pathway *in vivo*, a so-called biological pathway altering dose. This approach draws parallels between a chemical-associated perturbation of a pathway as observed in *in vitro* assays and a key event in the chemical's mode of action that may lead to an adverse health outcome. This approach offers an opportunity to not only compare alternatives with regards to the potential of human health hazard, but also take into account the quantitative and variability aspects of the underlying adverse effects.
3. Thomas *et al.* (2013) reasoned for a step-wise decision tree that incorporates *in vitro* assays, toxicokinetic modeling and short-term animal data into toxicity testing and risk assessment in an integrated fashion. Tier I of this approach informs the use of *in vitro* data in chemical alternative assessment. This phase uses *in vitro* assays to rank chemicals based on their relative selectivity in interacting with biological targets that have been associated with known toxicity outcomes and to identify the concentration at which these effects occur. Reverse toxicokinetic modeling and *in vitro*-to-*in vivo* extrapolation modeling (Wetmore, 2015) can then be used to convert *in vitro* concentrations into external dose for derivation of the point-of-departure values. The latter can be compared with human exposure data or estimates (Wambaugh *et al.*, 2013) to yield a margin of exposure.
4. The use of toxicokinetic information as a sufficient consideration in and of itself to enable rapid decisions on the risk of chemicals has been extended by Wambaugh *et al.* (2015) through the proposal for high-throughput physiologically based toxicokinetic model that predicts nonsteady-state

chemical concentration time-courses for a variety of exposure scenarios. The authors used this model to propose a 4-element framework for chemical toxicokinetic triage that can group chemicals into categories based on varying levels of confidence in predictions of exposure, an approach that may be useful for triaging chemicals for which high-throughput toxicokinetic predictions may be sufficient, and identifying those that may require additional experiments to collect appropriate data.

5. The U.S. EPA's Advancing the Next Generation of Risk Assessment program (Cote et al., 2016) considered options for how novel biological data and methods could better inform decision-making. New data and methods including transcriptomics, genomics, and proteomics; methods included molecular epidemiology and clinical studies, bioinformatic knowledge mining, pathway and network analyses, short-duration *in vivo* and *in vitro* bioassays, and quantitative structure activity relationship (QSAR) modeling were applied and evaluated for use in hazard identification and dose-response assessment. It was concluded that considerable uncertainties notwithstanding, application of new knowledge to risk assessment is warranted for the whole spectrum of decision contexts, from major scope assessments to prioritization and screening of very data limited chemicals.
6. The use of the *in vitro* and other novel data in the framework of the Adverse Outcome Pathway (AOP) is gaining interest with respect to decision-making applications (Edwards et al., 2016). The AOP concept has been proposed by the Organization for Economic Cooperation and Development (OECD) as a vehicle for linking the molecular screening and mechanistic toxicology data to *in vivo* adverse events of interest in human health and environmental assessments by describing a sequential progression from the chemical to the molecular initiating event, to the cellular, organ, organism and population response that underlies the *in vivo* outcome of interest (OECD, 2017). The overarching goal in using AOPs is to reduce the uncertainty in decision-making by identifying key intermediate events and quantitatively linking them to final adverse outcomes relevant to risk assessment. AOP framework provides transparency, allows for assessment of mechanistic probability, and enables hypothesis-based and decision context-relevant *in vitro* and *in silico* testing of a large number of chemicals. Conceptually, the AOP framework may therefore be very useful should a defined set of "adverse outcomes" be identified, eg, the types of adverse human health effects that are traditionally used in making "alternatives" or "design" chemical hazard assessment decisions.

USE OF NOVEL ASSESSMENT METHODOLOGIES-DERIVED DATA AND IN SILICO MODELS AS PRIMARY EVIDENCE

The National Toxicology Program Report on Carcinogens considers "studies on genotoxicity (ability to damage genes)" to be important evidence for the cancer hazard evaluation (National Toxicology Program, 2011). Many *in vitro* assays are available and widely used to test for the potential of a chemical to be genotoxic or mutagenic and a battery of well-defined tests is necessary for regulatory consideration of drugs and other chemicals (Doak et al., 2012). Indeed, many OECD guideline protocols for genotoxicity assessment have been established (Pfuhler et al., 2007) and a combination of the Ames test and *in vitro*

micronucleus assay can identify 78% of compounds known to be genotoxic *in vivo* (Kirkland et al., 2011). Standard OECD-approved assays are not meant for high-throughput testing and additional assays that can be used for screening of large chemical libraries are under evaluation (Fox et al., 2012; Yamamoto et al., 2011).

The utility of novel experimental tools to draw conclusion on the genotoxicity and potential tumorigenicity of a chemical has limitations (Benigni, 2013; Mahadevan et al., 2011); however, these assays can be used in a context of a comparative assessment. Similarly, the regulatory application of promising computational methods (eg, QSARs, Structure-Activity Relationships (SARs) and/or expert systems) for the evaluation of genotoxicity is growing, especially for impurities with limited to no information. A number of commercial and open-source statistically based and knowledge-based (expert system) tools are available and widely used (Fioravanzo et al., 2012). Computational models to assess the genotoxic potential of impurities in pharmaceutical products in the absence of experimental evidence are now incorporated into the International Community on Harmonization (ICH) M7 guidelines (<http://www.ich.org>; last assessed September 20, 2017). Although OECD guidelines exist for the validation and regulatory acceptance of (Q)SAR models, ICH M7 remains the only specific mention of the use of these *in silico* approaches as part of a regulatory guidance document.

The prohibition on placing animal-tested cosmetics on the market in Europe after 2013 created a tangible impetus to develop and validate alternative methodologies that can replace traditional *in vivo* tests (Adler et al., 2011). Not only were new methods developed, but a major effort was invested in developing reference databases and chemical lists to enable validation of the new methods to ensure that they can be used with confidence (Barroso et al., 2017), as well as *in silico* models for predictions of various aspects of dermal toxicity (Alves et al., 2015b). However, the opinions vary widely as to whether replacement of the current *in vivo* animal tests used for the safety assessment of cosmetic ingredients for skin sensitization and other outcomes is attainable. Although many experts agree that hazard-based classification of substances can be achieved based on *in vitro* tests and *in silico* models as primary data, estimations of the relative potency of a chemical and their toxicokinetic profiles would carry considerable uncertainty (Maxwell et al., 2014).

Finally, with the observed correlation between the onset of the proarrhythmic event Torsade's des Pointes and an observed prolongation of the QT interval, ie, the time from the beginning of the QRS complex to the end of the T wave of the electrocardiogram, most new pharmaceutical molecules are now required to undergo extensive testing to determine their effect on the QT interval. Part of this assessment includes the *in vitro* assessment of a chemical's ability to inhibit the I_{Kr} ionic current through the I_{Kr} channel as encoded by the hERG gene. This requirement is described more fully in the ICH S7B guidelines. However, it should be noted that inhibition of the hERG channel in isolation of a more comprehensive risk assessment is not indicative of being primary evidence for proarrhythmic risk.

USE OF IN VITRO DATA AND IN SILICO PREDICTIONS TO FILL DATA GAPS

Structure-based computational tools can be used to fill in certain data gaps. There has been a dramatic rise in the number of *in silico* models developed for human and environmental effects

as a result of large curated databases of toxicological information and increases in computational power. Health effects that now have *in silico* models developed for them include carcinogenicity (Contrera et al., 2007), hepatotoxicity (Greene et al., 2010), reproductive and developmental effects (Matthews et al., 2007; Wu et al., 2013), and skin sensitization (Alves et al., 2015a,b; Patlewicz et al., 2013).

Similarly, read-across can also be used to fill data gaps in chemical alternative assessment. Read-across is a process that uses the physicochemical properties and chemical structure of a chemical or group of chemicals that have been well studied toxicologically to infer the potential for toxicity of another similar chemical that lacks adequate toxicological data. For example, for almost 2 decades, some organizations have been qualitatively treating the toxicity of chemicals that bear close structural similarity to dioxin 2,3,7,8-tetrachlorodibenzo-p-dioxin and that meet certain criteria of being “dioxin-like” as producing the same type of toxicity as dioxin (Van den Berg et al., 2006).

With the publication of the European Union Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) regulations read-across approaches have received more attention because they can help satisfy the information requirements under these regulations. The general concept of read-across has now been accepted by the European Chemicals Agency (ECHA) and Member State Authorities (Patlewicz et al., 2013). ECHA has recently published a framework by which it evaluates read-across submissions under REACH (European Chemicals Agency, 2015).

However, in conducting read-across or other types of chemical structure-based assessments it is important to include a consideration of the physicochemical properties of the comparator compounds. The role of hydrophobicity and other physicochemical properties of a molecule can influence its bioavailability, metabolic stability, tissue distribution, rate of clearance, and ultimate toxicity. These associations have been detailed in the publications by Lipinski et al. (Lipinski, 2016) who developed the Lipinski's Rule of 5. Properties such as a compound's octanol/water partition coefficient (LogP) have a profound influence on drug potency, pharmacokinetics and toxicity (Cronin, 2006; Leeson and Springthorpe, 2007). A cross-industry review (Waring et al., 2015) of pharmaceutical successes and failures showed there to be a small but statistically significant difference in the LogP values for those compounds that were terminated for clinical safety and those that went on to later stages of the development process.

High-throughput *in vitro* data can also be used to fill certain primary data gaps (eg, data gaps for particular health endpoints) or for replacing some animal tests. For example, *in vitro* assays of nuclear receptors, cytochrome P450 enzymes, G protein-coupled receptors, and cell signaling pathways were informative in predicting rodent reproductive and developmental toxicants (Martin et al., 2011; Sipes et al., 2011). It was demonstrated recently that results from 18 estrogen receptor ToxCast high-throughput screening assays and a computational model that can discriminate bioactivity from assay-specific interference and cytotoxicity can substitute for the rat uterotherphic assay (Browne et al., 2015; Judson et al., 2015).

Availability of novel *in vitro* data and *in silico* predictions also requires new approaches to integrative data analysis, where diverse information is presented and assessed in a comprehensive manner. The exponentially increasing quantities and sources of new toxicity testing-derived information must be anchored to well-studied historical toxicity data. The methods

used for integrative analysis must be flexible and easily communicated to all stakeholders, including regulators, and there is a need to automate a read-across process to combine both chemical and biological factors, while keeping the process transparent for expert interrogation (Berggren et al., 2015). To address these challenges, several approaches to visualize multidimensional information have been developed (Figure 1). For example, Low et al. (2013) proposed a hazard classification and visualization method (Figure 1A) that draws upon both chemical structural similarity and comparisons of biological responses to chemicals evaluated in multiple *in vitro* assays (“biological” similarity). Specifically, this approach interpolates each compound's potential for toxicity from those of chemical and biological “analogues” where similarities are determined by the Tanimoto coefficient. The ToxPi framework (Reif et al., 2010, 2013) provides a visual translation of the integration and weighting scheme for chemical and substance prioritization, while providing statistical metrics of robustness and uncertainty associated with the decision (Figure 1B).

The ToxPi has been used as data-integrating approach (Grimm et al., 2016) of grouping chemically complex substances (products of petroleum refining) to demonstrate how novel data streams could be used as primary data in human health assessments or to fill data gaps across a broad range of domains, including health. Indeed, read across using chemical-biological data integration has been an approach aimed at bridging novel data streams and chemical biology and structure-activity modeling (Greene and Song, 2011; Low et al., 2013, 2014). Interpretation of such multidimensional data sets is not limited to biological and/or chemical property-derived groupings for regulatory applications, but is also informative for mechanistic toxicity evaluations.

THE USE OF *IN VITRO* DATA TO SCREEN OUT POSSIBLE UNINTENDED CONSEQUENCES

One goal of alternatives assessment is to “reduce unwanted consequences” of chemical alternatives. This goal is also shared by the pharmaceutical industry especially in the drug safety evaluation step. The pharmaceutical industry routinely uses *in vitro* tests and high-throughput screening for evaluation of drug candidates (MacDonald and Robertson, 2009). *In vitro* high-throughput screening and toxicogenomics appear promising in the screening of data-poor chemical alternatives for modes of action that may contribute to hazard identification and dose-response assessments.

One area in which *in vitro* data may be useful is in the evaluation of human variability. Human variability underlies differences in the degrees and ways in which people respond to environmental chemicals and addressing these differences is a key consideration in human health risk assessments for chemicals (National Research Council, 2009). Assessments of alternatives may take advantage of novel *in vitro* data to both characterize and quantify variability (Eduati et al., 2015; Zeise et al., 2013). The utility of *in vitro* population-based models to toxicology, especially for exploring the extent and nature of genetic components of interindividual variability in pharmacodynamics, was recently demonstrated (Abdo et al., 2015a,b; Lock et al., 2012; O'Shea et al., 2011). Quantitative high-throughput screening for cytotoxicity endpoints can be used for variability assessment in concentration-response, to explore the potential genetic determinants of the variability, and to derive mode of

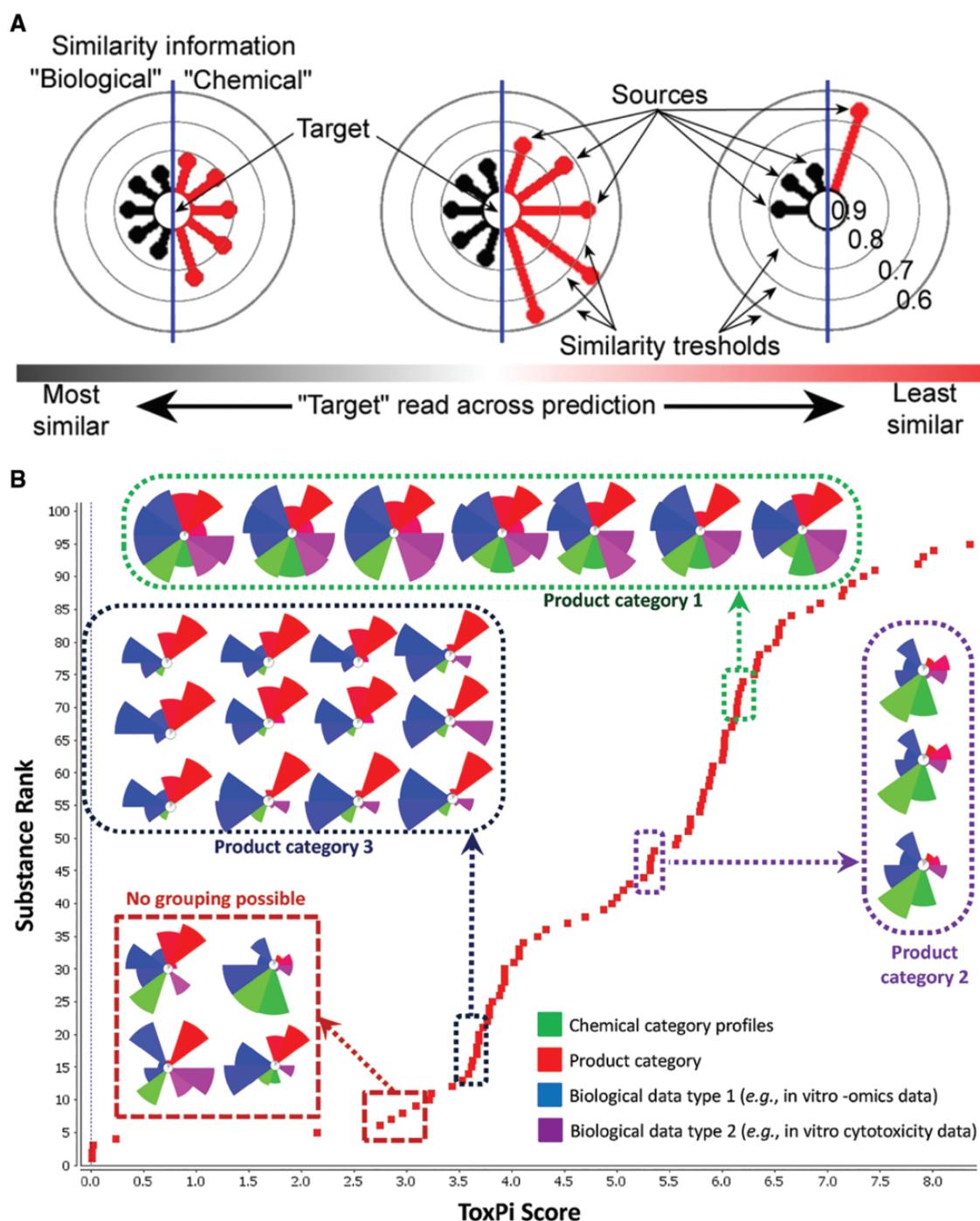


Figure 1. Data integration and visualization approaches applicable to novel toxicity data streams. A, Analog read across through multi-dimensional analysis of chemical and biological data (Low *et al.*, 2013). B, ToxPi-based category read-across (Reif *et al.*, 2010, 2013).

action hypotheses for follow-up analyses. The development and use of these and other types of *in vitro* assays, including organotypic and microphysiological culture systems, would be further informed by quantitative comparisons of the interindividual variability measured *in vitro* with observable human variability *in vivo*. Interindividual variability in response to chemicals in *in vitro* assays could also be compared with *in vivo* human pharmacodynamics variability data.

Use of human genetic data to identify populations that are more susceptible to certain types of toxicants is an area of

research that is in its infancy. However, with advances in the area personalized medicine (Hamburg and Collins, 2010) comes an increased knowledge of how chemicals interact with a biological system and the potential for unintended consequences from exposure to chemicals. Indeed, the initial results of modeling the linkages between the individual's genotype and variability in toxicity from environmental chemicals and drugs are encouraging, even if the accuracy of the predictions needs to be improved with additional data (Eduati *et al.*, 2015).

LIMITATIONS OF USING IN SILICO AND IN VITRO DATA IN ALTERNATIVES ASSESSMENTS

The promise of the novel assays and computational methods is difficult to underestimate; however, it is still not clear what is the predictive power or classification accuracy of the *in vitro* assays. It has been stated that the data from *in vitro* assays used in isolation may not be superior to predictions from that of QSAR models and aggregating the assays based on genes or pathways may even lead to reduced predictive performance (Thomas *et al.*, 2012). Potential bias in the predictions from the classification models have been noted, albeit this point is still a subject of controversy (Dix *et al.*, 2012; Knudsen *et al.*, 2013). Thus, the current high-throughput *in vitro* assays may still have only limited applicability for chemical alternatives analysis and risk assessment.

It is evident that the majority of environmental chemicals being tested in large toxicity screening efforts likely exert toxicity via nonselective interactions with cellular macromolecules (Judson *et al.*, 2016; Thomas *et al.*, 2013). Because most high-throughput assays incorporate extensive concentration ranges, it is frequently observed that many biological targets are “engaged” at, or near, cytotoxic concentrations (Judson *et al.*, 2016; Martin *et al.*, 2010). Thus, one use of high-throughput assay-derived information may be for separating chemicals those that are toxic via nonselective interactions and those that act through selective interactions (eg, receptor-mediated effects). Such separation according to relative “selectivity” may allow for a more precise definition of the mode of action and the point of departure for the agents that exhibit high selectivity towards particular molecular targets in the screening battery, while for the nonselective ones the elucidation of the exact mechanistic events may be more challenging. In the context of chemical alternatives analysis, should alternatives under consideration exhibit varying levels of “selectivity”, a compound with higher “selectivity” may be considered by the assessor as lower risk for additional “off target” effects and thus be assigned a higher relative rank.

Similarly, there has been a significant increase in development of *in silico* models and the acceptance of these into regulatory guidelines as primary evidence of human health effects has set the scene for further development in this area. However, the acceptance and implementation of *in silico* models remains highly variable and dependent on the health effect being predicted. This is partly due to the predictive performance of the models for some health effects fall short of that required for practical applications, but is also due in part to the fact that most approaches only provide a binary prediction, ie the chemical will either cause the adverse effect or it won't. The inability to predict a plasma concentration that would be expected to illicit the toxicity ultimately limits the utility and incorporation of *in silico* systems as a part of a risk assessment process or an innovative design process as they often cannot differentiate between closely related structures when there is little or no safety information available for comparison.

This limitation of *in silico* QSAR models has in part led to the increased use of approaches like read-across. However, read-across approaches are not without their problems and it is often necessary to understand the chemical drivers of the observed toxicity in order to determine what chemicals are to be considered as similar enough in order to infer activity. There are numerous examples where single substitution differences can have a profound effect on a chemicals toxicity profile, eg, the

neurotoxicity of n-heptane versus n-hexane versus n-pentane (Takeuchi *et al.*, 1980).

CONCLUSIONS

In the past decade, the state of the science and decision-making in scenarios when little to no data may be available to draw conclusions about which substances represent a “safer” option in green chemistry applications has been transformed. Options exist based on the context of the decision and the availability of resources, and range from the approaches that require little new information to be collected, to extensive screening programs in cells, organotypic models, and whole animals, as was recently detailed by another National Academies report (National Academies of Sciences, 2017). It is clear that there is no “one size fits all” solution and multiple data streams need to be weighed in making a decision. Moreover, the overall level of familiarity of the decision-makers and scientists alike with new assessment methodologies, their validity, value and limitations is evolving. Thus, while the “impact” of the new developments in toxicology on the field of green chemistry is great already, it is premature to conclude that the data from new assessment methodologies have been widely accepted yet.

FUNDING

This publication was made possible, in part, by NIH grant P42 ES027704. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. Further, the NIH does not endorse the purchase of any commercial products or services mentioned in the publication.

ACKNOWLEDGMENTS

The authors acknowledge members of the National Research Council Committee on Design and Evaluation of Safer Chemical Substitutions—A Framework to Inform Government and Industry Decisions and NRC staff for valuable discussions of the topics covered in this article.

REFERENCES

- Abdo, N., Wetmore, B. A., Chappell, G. A., Shea, D., Wright, F. A., and Rusyn, I. (2015a). *In vitro* screening for population variability in toxicity of pesticide-containing mixtures. *Environ. Int.* **85**, 147–155.
- Abdo, N., Xia, M., Brown, C. C., Kosyk, O., Huang, R., Sakamuru, S., Zhou, Y. H., Jack, J. R., Gallins, P., Xia, K., *et al.* (2015b). Population-based *in vitro* hazard and concentration-response assessment of chemicals: The 1000 genomes high-throughput screening study. *Environ. Health Perspect.* **123**, 458–466.
- Adler, S., Basketter, D., Creton, S., Pelkonen, O., van Benthem, J., Zuang, V., Andersen, K. E., Angers-Loustau, A., Aptula, A., Bal-Price, A., *et al.* (2011). Alternative (non-animal) methods for cosmetics testing: Current status and future prospects-2010. *Arch. Toxicol.* **85**, 367–485.
- Alves, V. M., Muratov, E., Fourches, D., Strickland, J., Kleinstreuer, N., Andrade, C. H., and Tropsha, A. (2015a). Predicting chemically-induced skin reactions. Part I: QSAR models of skin sensitization and their application to identify

- potentially hazardous compounds. *Toxicol. Appl. Pharmacol.* **284**, 262–272.
- Alves, V. M., Muratov, E., Fourches, D., Strickland, J., Kleinstreuer, N., Andrade, C. H., and Tropsha, A. (2015b). Predicting chemically-induced skin reactions. Part II: QSAR models of skin permeability and the relationships between skin permeability and skin sensitization. *Toxicol. Appl. Pharmacol.* **284**, 273–280.
- Andersen, M. E., and Krewski, D. (2009). Toxicity testing in the 21st century: Bringing the vision to life. *Toxicol. Sci.* **107**, 324–330.
- Andersen, M. E., and Krewski, D. (2010). The vision of toxicity testing in the 21st century: Moving from discussion to action. *Toxicol. Sci.* **117**, 17–24.
- Barroso, J., Pfannenbecker, U., Adriaens, E., Alepee, N., Cluzel, M., De Smedt, A., Hibatallah, J., Klaric, M., Mewes, K. R., Millet, M., et al. (2017). Cosmetics Europe compilation of historical serious eye damage/eye irritation in vivo data analysed by drivers of classification to support the selection of chemicals for development and evaluation of alternative methods/strategies: The Draize eye test Reference Database (DRD). *Arch. Toxicol.* **91**, 521–547.
- Benigni, R. (2013). Evaluation of the toxicity forecasting capability of EPA's ToxCast Phase I data: Can ToxCast in vitro assays predict carcinogenicity? *J. Environ. Sci. Health C Environ. Carcinog. Ecotoxicol. Rev.* **31**, 201–212.
- Berggren, E., Amcoff, P., Benigni, R., Blackburn, K., Carney, E., Cronin, M., Deluyker, H., Gautier, F., Judson, R. S., Kass, G. E., et al. (2015). Chemical safety assessment using read-across: Assessing the use of novel testing methods to strengthen the evidence base for decision making. *Environ. Health Perspect.* **123**, 1232–1240.
- Bhatia, S., Schultz, T., Roberts, D., Shen, J., Kromidas, L., and Marie Api, A. (2015). Comparison of Cramer classification between Toxtree, the OECD QSAR Toolbox and expert judgment. *Regul. Toxicol. Pharmacol.* **71**, 52–62.
- Browne, P., Judson, R. S., Casey, W. M., Kleinstreuer, N. C., and Thomas, R. S. (2015). Screening chemicals for estrogen receptor bioactivity using a computational model. *Environ. Sci. Technol.* **49**, 8804–8814.
- Collins, F. S., Gray, G. M., and Bucher, J. R. (2008). Toxicology. Transforming environmental health protection. *Science* **319**, 906–907.
- Contrera, J. F., Kruhlak, N. L., Matthews, E. J., and Benz, R. D. (2007). Comparison of MC4PC and MDL-QSAR rodent carcinogenicity predictions and the enhancement of predictive performance by combining QSAR models. *Regul. Toxicol. Pharmacol.* **49**, 172–182.
- Cote, I., Andersen, M. E., Ankley, G. T., Barone, S., Birnbaum, L. S., Boekelheide, K., Bois, F. Y., Burgoon, L. D., Chiu, W. A., Crawford-Brown, D., et al. (2016). The Next generation of risk assessment multi-year study-highlights of findings, applications to risk assessment, and future directions. *Environ. Health Perspect.* **124**, 1671–1682.
- Cramer, G. M., Ford, R. A., and Hall, R. L. (1978). Estimation of toxic hazard—a decision tree approach. *Food Cosmet. Toxicol.* **16**, 255–276.
- Cronin, M. T. (2006). The role of hydrophobicity in toxicity prediction. *Curr. Comput. Aid. Drug Des.* **2**, 405–413.
- Crump, K. S., Chen, C., and Louis, T. A. (2010). The future use of in vitro data in risk assessment to set human exposure standards: Challenging problems and familiar solutions. *Environ. Health Perspect.* **118**, 1350–1354.
- Dix, D. J., Houck, K. A., Judson, R. S., Kleinstreuer, N. C., Knudsen, T. B., Martin, M. T., Reif, D. M., Richard, A. M., Shah, I., Sipes, N. S., et al. (2012). Incorporating biological, chemical, and toxicological knowledge into predictive models of toxicity. *Toxicol. Sci.* **130**, 440–441. author reply 442–443.
- Doak, S. H., Manshian, B., Jenkins, G. J., and Singh, N. (2012). In vitro genotoxicity testing strategy for nanomaterials and the adaptation of current OECD guidelines. *Mutat. Res.* **745**, 104–111.
- Eduati, F., Mangravite, L. M., Wang, T., Tang, H., Bare, J. C., Huang, R., Norman, T., Kellen, M., Menden, M. P., Yang, J., et al. (2015). Prediction of human population responses to toxic compounds by a collaborative competition. *Nat. Biotechnol.* **33**, 933–940.
- Edwards, S. W., Tan, Y. M., Villeneuve, D. L., Meek, M. E., and McQueen, C. A. (2016). Adverse outcome pathways-organizing toxicological information to improve decision making. *J. Pharmacol. Exp. Ther.* **356**, 170–181.
- Enoch, S. J., Cronin, M. T., Schultz, T. W., and Madden, J. C. (2008). Quantitative and mechanistic read across for predicting the skin sensitization potential of alkenes acting via Michael addition. *Chem. Res. Toxicol.* **21**, 513–520.
- European Chemicals Agency. (2017). Read-Across Assessment Framework (RAFF). Document ECHA-17-R-01-EN. European Chemicals Agency. Helsinki, Finland.
- Fioravanzo, E., Bassan, A., Pavan, M., Mostrag-Szlichtyng, A., and Worth, A. P. (2012). Role of in silico genotoxicity tools in the regulatory assessment of pharmaceutical impurities. *SAR QSAR Environ. Res.* **23**, 257–277.
- Fox, J. T., Sakamuru, S., Huang, R., Teneva, N., Simmons, S. O., Xia, M., Tice, R. R., Austin, C. P., and Myung, K. (2012). High-throughput genotoxicity assay identifies antioxidants as inducers of DNA damage response and cell death. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 5423–5428.
- Greene, N., Fisk, L., Naven, R. T., Note, R. R., Patel, M. L., and Pelletier, D. J. (2010). Developing structure-activity relationships for the prediction of hepatotoxicity. *Chem. Res. Toxicol.* **23**, 1215–1222.
- Greene, N., and Song, M. (2011). Predicting in vivo safety characteristics using physiochemical properties and in vitro assays. *Future Med. Chem.* **3**, 1503–1511.
- Grimm, F. A., Iwata, Y., Sirenko, O., Chappell, G. A., Wright, F. A., Reif, D. M., Braisted, J., Gerhold, D. L., Yeakley, J. M., Shepard, P., et al. (2016). A chemical-biological similarity-based grouping of complex substances as a prototype approach for evaluating chemical alternatives. *Green Chem.* **18**, 4407–4419.
- Grimm, F. A., Russell, W. K., Luo, Y. S., Iwata, Y., Chiu, W. A., Roy, T., Boogaard, P. J., Ketelslegers, H. B., and Rusyn, I. (2017). Grouping of petroleum substances as example UVCBs by ion mobility-mass spectrometry to enable chemical composition-based read-across. *Environ. Sci. Technol.* **51**, 7197–7207.
- Hamburg, M. A., and Collins, F. S. (2010). The path to personalized medicine. *N. Engl. J. Med.* **363**, 301–304.
- Hewitt, M., Ellison, C. M., Enoch, S. J., Madden, J. C., and Cronin, M. T. (2010). Integrating (Q)SAR models, expert systems and read-across approaches for the prediction of developmental toxicity. *Reprod. Toxicol.* **30**, 147–160.
- Judson, R., Houck, K., Martin, M., Richard, A. M., Knudsen, T. B., Shah, I., Little, S., Wambaugh, J., Woodrow Setzer, R., Kothya, P., et al. (2016). Editor's highlight: Analysis of the effects of cell stress and cytotoxicity on in vitro assay activity across a diverse chemical and assay space. *Toxicol. Sci.* **152**, 323–339.

- Judson, R. S., Kavlock, R. J., Setzer, R. W., Hubal, E. A., Martin, M. T., Knudsen, T. B., Houck, K. A., Thomas, R. S., Wetmore, B. A., and Dix, D. J. (2011). Estimating toxicity-related biological pathway altering doses for high-throughput chemical risk assessment. *Chem. Res. Toxicol.* **24**, 451–462.
- Judson, R. S., Magpantay, F. M., Chickarmane, V., Haskell, C., Tania, N., Taylor, J., Xia, M., Huang, R., Rotroff, D. M., Filer, D. L., et al. (2015). Integrated model of chemical perturbations of a biological pathway using 18 in vitro high-throughput screening assays for the estrogen receptor. *Toxicol. Sci.* **148**, 137–154.
- Kavlock, R., Chandler, K., Houck, K., Hunter, S., Judson, R., Kleinstreuer, N., Knudsen, T., Martin, M., Padilla, S., Reif, D., et al. (2012). Update on EPA's ToxCast program: Providing high throughput decision support tools for chemical risk management. *Chem. Res. Toxicol.* **25**, 1287–1302.
- Kirkland, D., Reeve, L., Gatehouse, D., and Vanparys, P. (2011). A core in vitro genotoxicity battery comprising the Ames test plus the in vitro micronucleus test is sufficient to detect rodent carcinogens and in vivo genotoxins. *Mutat. Res.* **721**, 27–73.
- Knudsen, T., Martin, M., Chandler, K., Kleinstreuer, N., Judson, R., and Sipes, N. (2013). Predictive models and computational toxicology. *Methods Mol. Biol.* **947**, 343–374.
- Krewski, D., Westphal, M., Al Zoughool, M., Croteau, M. C., and Andersen, M. E. (2011). New directions in toxicity testing. *Annu. Rev. Public Health* **32**, 161–178.
- Leeson, P. D., and Springthorpe, B. (2007). The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Discov.* **6**, 881–890.
- Lipinski, C. A. (2016). Rule of five in 2015 and beyond: Target and ligand structural limitations, ligand chemistry structure and drug discovery project decisions. *Adv. Drug Deliv. Rev.* **101**, 34–41.
- Lock, E. F., Abdo, N., Huang, R., Xia, M., Kosyk, O., O'Shea, S. H., Zhou, Y. H., Sedykh, A., Tropsha, A., Austin, C. P., et al. (2012). Quantitative high-throughput screening for chemical toxicity in a population-based in vitro model. *Toxicol. Sci.* **126**, 578–588.
- Low, Y., Sedykh, A., Fourches, D., Golbraikh, A., Whelan, M., Rusyn, I., and Tropsha, A. (2013). Integrative chemical-biological read-across approach for chemical hazard classification. *Chem. Res. Toxicol.* **26**, 1199–1208.
- Low, Y. S., Sedykh, A. Y., Rusyn, I., and Tropsha, A. (2014). Integrative approaches for predicting in vivo effects of chemicals from their structural descriptors and the results of short-term biological assays. *Curr. Top. Med. Chem.* **14**, 1356–1364.
- MacDonald, J. S., and Robertson, R. T. (2009). Toxicity testing in the 21st century: A view from the pharmaceutical industry. *Toxicol. Sci.* **110**, 40–46.
- Mahadevan, B., Snyder, R. D., Waters, M. D., Benz, R. D., Kemper, R. A., Tice, R. R., and Richard, A. M. (2011). Genetic toxicology in the 21st century: Reflections and future directions. *Environ. Mol. Mutagen.* **52**, 339–354.
- Martin, M. T., Dix, D. J., Judson, R. S., Kavlock, R. J., Reif, D. M., Richard, A. M., Rotroff, D. M., Romanov, S., Medvedev, A., Poltoratskaya, N., et al. (2010). Impact of environmental chemicals on key transcription regulators and correlation to toxicity end points within EPA's ToxCast program. *Chem. Res. Toxicol.* **23**, 578–590.
- Martin, M. T., Knudsen, T. B., Reif, D. M., Houck, K. A., Judson, R. S., Kavlock, R. J., and Dix, D. J. (2011). Predictive model of rat reproductive toxicity from ToxCast high throughput screening. *Biol. Reprod.* **85**, 327–339.
- Matthews, E. J., Kruhlak, N. L., Daniel Benz, R., Ivanov, J., Klopman, G., and Contrera, J. F. (2007). A comprehensive model for reproductive and developmental toxicity hazard identification: II. Construction of QSAR models to predict activities of untested chemicals. *Regul. Toxicol. Pharmacol.* **47**, 136–155.
- Maxwell, G., MacKay, C., Cubberley, R., Davies, M., Gellatly, N., Glavin, S., Guin, T., Jacquilleot, S., Moore, C., Pendlington, R., et al. (2014). Applying the skin sensitisation adverse outcome pathway (AOP) to quantitative risk assessment. *Toxicol. In Vitro* **28**, 8–12.
- National Academies of Sciences. (2017). *Using 21st Century Science to Improve Risk-Related Evaluations*. Washington, DC: The National Academies Press.
- National Research Council. (2007). *Toxicity Testing in the 21st Century: A Vision and A Strategy*. The National Academies Press, Washington, DC.
- National Research Council. (2009). *Science and Decisions: Advancing Risk Assessment*. National Academies Press, Washington, DC.
- National Research Council. (2014). *A Framework to Guide Selection of Chemical Alternatives*. The National Academies Press, Washington, DC.
- National Toxicology Program. (2011). *12th Report on Carcinogens*. National Institute of Environmental Health Sciences, Research Triangle Park, NC.
- O'Shea, S. H., Schwarz, J., Kosyk, O., Ross, P. K., Ha, M. J., Wright, F. A., and Rusyn, I. (2011). In vitro screening for population variability in chemical toxicity. *Toxicol. Sci.* **119**, 398–407.
- OECD. (2017) Revised Guidance Document on Developing and Assessing Adverse Outcome Pathways. Document ENV/JM/MONO(2013)6. Environment Directorate of the Organization for Economic Cooperation and Development. Paris, France.
- Patlewicz, G., Ball, N., Booth, E. D., Hulzebos, E., Zvinavashe, E., and Hennes, C. (2013). Use of category approaches, read-across and (Q)SAR: General considerations. *Regul. Toxicol. Pharmacol.* **67**, 1–12.
- Pfuhler, S., Albertini, S., Fautz, R., Herbold, B., Madle, S., Utesch, D., Poth, A., and Gesellschaft fuer, U.-M. (2007). Genetic toxicity assessment: Employing the best science for human safety evaluation part IV: Recommendation of a working group of the Gesellschaft fuer Umwelt-Mutationsforschung (GUM) for a simple and straightforward approach to genotoxicity testing. *Toxicol. Sci.* **97**, 237–240.
- Reif, D. M., Martin, M. T., Tan, S. W., Houck, K. A., Judson, R. S., Richard, A. M., Knudsen, T. B., Dix, D. J., and Kavlock, R. J. (2010). Endocrine profiling and prioritization of environmental chemicals using ToxCast data. *Environ. Health Perspect.* **118**, 1714–1720.
- Reif, D. M., Sypa, M., Lock, E. F., Wright, F. A., Wilson, A., Cathey, T., Judson, R. R., and Rusyn, I. (2013). ToxPi GUI: An interactive visualization tool for transparent integration of data from diverse sources of evidence. *Bioinformatics* **29**, 402–403.
- Roberts, D. W., Aptula, A., Schultz, T. W., Shen, J., Api, A. M., Bhatia, S., and Kromidas, L. (2015). A practical guidance for Cramer class determination. *Regul. Toxicol. Pharmacol.* **73**, 971–984.
- Rusyn, I., and Daston, G. P. (2010). Computational toxicology: Realizing the promise of the toxicity testing in the 21st century. *Environ. Health Perspect.* **118**, 1047–1050.
- Rusyn, I., Sedykh, A., Low, Y., Guyton, K. Z., and Tropsha, A. (2012). Predictive modeling of chemical hazard by integrating

- numerical descriptors of chemical structures and short-term toxicity assay data. *Toxicol. Sci.* **127**, 1–9.
- Sipes, N. S., Martin, M. T., Reif, D. M., Kleinstreuer, N. C., Judson, R. S., Singh, A. V., Chandler, K. J., Dix, D. J., Kavlock, R. J., and Knudsen, T. B. (2011). Predictive models of prenatal developmental toxicity from ToxCast high-throughput screening data. *Toxicol. Sci.* **124**, 109–127.
- Takeuchi, Y., Ono, Y., Hisanaga, N., Kitoh, J., and Sugiura, Y. (1980). A comparative study on the neurotoxicity of n-pentane, n-hexane, and n-heptane in the rat. *Br. J. Ind. Med.* **37**, 241–247.
- Thomas, R. S., Black, M. B., Li, L., Healy, E., Chu, T. M., Bao, W., Andersen, M. E., and Wolfinger, R. D. (2012). A comprehensive statistical analysis of predicting in vivo hazard using high-throughput in vitro screening. *Toxicol. Sci.* **128**, 398–417.
- Thomas, R. S., Philbert, M. A., Auerbach, S. S., Wetmore, B. A., Devito, M. J., Cote, I., Rowlands, J. C., Whelan, M. P., Hays, S. M., Andersen, M. E., et al. (2013). Incorporating new technologies into toxicity testing and risk assessment: Moving from 21st century vision to a data-driven framework. *Toxicol. Sci.* **136**, 4–18.
- Tice, R. R., Austin, C. P., Kavlock, R. J., and Bucher, J. R. (2013). Improving the human hazard characterization of chemicals: A Tox21 update. *Environ. Health Perspect.* **121**, 756–765.
- Gluczkiewicz, I., Buist, H. E., Martin, M. T., Mangelsdorf, I., and Escher, S. E. (2011). Improvement of the Cramer classification for oral exposure using the database TTC RepDose—a strategy description. *Regul. Toxicol. Pharmacol.* **61**, 340–350.
- Gluczkiewicz, I., Kuhne, R., Ebert, R. U., Batke, M., Schuurmann, G., Mangelsdorf, I., and Escher, S. E. (2016). Inhalation TTC values: A new integrative grouping approach considering structural, toxicological and mechanistic features. *Regul. Toxicol. Pharmacol.* **78**, 8–23.
- Van den Berg, M., Birnbaum, L. S., Denison, M., De Vito, M., Farland, W., Feeley, M., Fiedler, H., Hakansson, H., Hanberg, A., Haws, L., et al. (2006). The 2005 World Health Organization reevaluation of human and Mammalian toxic equivalency factors for dioxins and dioxin-like compounds. *Toxicol. Sci.* **93**, 223–241.
- Voutchkova, A. M., Osimitz, T. G., and Anastas, P. T. (2010). Toward a comprehensive molecular design framework for reduced hazard. *Chem. Rev.* **110**, 5845–5882.
- Wambaugh, J. F., Setzer, R. W., Reif, D. M., Gangwal, S., Mitchell-Blackwood, J., Arnot, J. A., Joliet, O., Frame, A., Rabinowitz, J., Knudsen, T. B., et al. (2013). High-throughput models for exposure-based chemical prioritization in the ExpoCast project. *Environ. Sci. Technol.* **47**, 8479–8488.
- Wambaugh, J. F., Wetmore, B. A., Pearce, R., Strobe, C., Goldsmith, R., Sluka, J. P., Sedykh, A., Tropsha, A., Bosgra, S., Shah, I., et al. (2015). Toxicokinetic triage for environmental chemicals. *Toxicol. Sci.* **147**, 55–67.
- Waring, M. J., Arrowsmith, J., Leach, A. R., Leeson, P. D., Mandrell, S., Owen, R. M., Pairaudeau, G., Pennie, W. D., Pickett, S. D., Wang, J., et al. (2015). An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat. Rev. Drug Discov.* **14**, 475–486.
- Wetmore, B. A. (2015). Quantitative in vitro-to-in vivo extrapolation in a high-throughput environment. *Toxicology* **332**, 94–101.
- Wu, S., Fisher, J., Naciff, J., Laufersweiler, M., Lester, C., Daston, G., and Blackburn, K. (2013). Framework for identifying chemicals with structural features associated with the potential to act as developmental or reproductive toxicants. *Chem. Res. Toxicol.* **26**, 1840–1861.
- Yamamoto, K. N., Hirota, K., Kono, K., Takeda, S., Sakamuru, S., Xia, M., Huang, R., Austin, C. P., Witt, K. L., and Tice, R. R. (2011). Characterization of environmental chemicals with potential for DNA damage using isogenic DNA repair-deficient chicken DT40 cell lines. *Environ. Mol. Mutagen.* **52**, 547–561.
- Zeise, L., Bois, F. Y., Chiu, W. A., Hattis, D., Rusyn, I., and Guyton, K. Z. (2013). Addressing human variability in next-generation human health risk assessments of environmental chemicals. *Environ. Health Perspect.* **121**, 23–31.